

Alignment methods for biased multicanonical sampling

Michael Reimer,^{1,*} Ahmed Awadalla,¹ David Yevick,¹ and Tao Lu²

¹*Department of Physics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*

²*Department of Physics, California Institute of Technology, Pasadena, California 91125, USA*

*Corresponding author: mareimer@sympatico.ca

Received January 24, 2007; accepted March 1, 2007;
posted March 23, 2007 (Doc. ID 79311); published July 11, 2007

The efficiency of the multicanonical procedure can be significantly improved by applying an additional bias to the numerically generated sample space. However, results obtained by biasing in different sampling regions cannot in general be accurately combined, since their relative normalization coefficient is not known precisely. We demonstrate that for overlapping biasing regions a simple iterative procedure can be employed to determine the required coefficients. © 2007 Optical Society of America

OCIS codes: 060.0060, 060.2330, 060.2400, 260.5430.

1. INTRODUCTION

Communication system studies normally involve the estimation of small error probabilities that are associated with physically unlikely system configurations [1,2]. Numerical system simulations have therefore recently employed the multicanonical method [3–6] to increase the likelihood of physically interesting sampling events. In our initial studies of the probability density function (pdf) of the differential group delay (DGD) of an optical fiber [7–9], we found that multicanonical procedures reduced the number of samples required to predict or measure the low-probability region of the pdf by orders of magnitude. However, the increase in computational efficiency, while large, can still be insufficient for experimental applications [10–12] in which far fewer system configurations can be sampled compared with numerical simulations. We accordingly proposed several improvements to multicanonical sampling [13,14], including biased multicanonical approaches that concentrate the statistical samples into restricted regions of the space of system observables. With appropriate bias functions [15], we performed measurements of the pdf of the DGD to an unparalleled level of precision with standard equipment [12].

However, if our method is applied repeatedly to the same problem but with the sample space biased in different regions, the associated histograms cannot be combined to predict the global system behavior, since their relative normalization is undetermined. Here we present a procedure for calculating the required normalization constants when the sampling regions of successive biased multicanonical iterations (or separate multicanonical calculations) overlap. In particular, the relative normalization constants are obtained from an iterative procedure in which the converged solution minimizes the mean square error (MSE) of the pdf estimate in the overlapping regions. Our method yields a substantial increase in computational efficiency with modest additional programming effort.

Methods similar to ours have been previously applied to the Monte Carlo method and to the canonical distribution [16] generated by the Metropolis algorithm. These include successive umbrella sampling [17–19], bridge sampling [20], importance sampling, and the weighted histogram analysis method of [21–23]. However, the procedure of this paper is, to our knowledge, the first implementation of such techniques to the multicanonical ensemble [14,24–26].

2. MULTICANONICAL SAMPLING

The multicanonical method is an iterative, biased statistical sampling method that presupposes a complete absence of prior information about the system properties. The procedure generates the value of any function of system variables, $p'(\vec{E})$, weighted by its probability of occurrence, which is collectively denoted by $p(\vec{E})$. The $\vec{E}(\vec{\alpha})$ are a vector of N_E system observables that are dependent on N_α stochastic parameters, $\vec{\alpha}$. In the calculations of this paper, the observables correspond to the DGD of an optical fiber system, the randomly varying parameters, $\vec{\alpha}$, are the angles of various polarization controllers [11,12], and $p(\vec{E})$ is the pdf associated with finding the system in a given state of the observable. However, it should be emphasized that the choice of these quantities is effectively arbitrary (in the more general case, which is discussed and illustrated in detail in [14], the histograms below must be updated by a value different from unity).

To implement the multicanonical method, the physically relevant region of the output values (solution space) \vec{E} is divided into N histogram bins with volume ΔV centered at \vec{E}_i with $i=1, 2, \dots, N$. A histogram that here corresponds to the current estimate of the (unnormalized) pdf, which we denote $p_0(\vec{E})$, and a second histogram, $H_1(\vec{E})$, that stores temporary values, are first set to unity, consistent with an absence of prior information. We then gener-

ate a set of random system variables, $\vec{\alpha}^{cur}$, and compute the associated observables \vec{E}^{cur} . A transition is then generated from $\vec{\alpha}^{new} = \vec{\alpha}^{cur} + \Delta\vec{\alpha}$, where $\Delta\vec{\alpha}$ represents a small random perturbation (which can be chosen from an effectively arbitrary probability distribution, as verified in [14]), leading to new observables \vec{E}^{new} . This transition is accepted with probability $\min[1, p_0(\vec{E}^{cur})/p_0(\vec{E}^{new})]$, in which case the system variables $\vec{\alpha}^{cur}$ are replaced by $\vec{\alpha}^{new}$; otherwise, a self-transition occurs, and $\vec{\alpha}^{new}$ is set to $\vec{\alpha}^{cur}$ for the following step [6]. In either case, the value of the histogram bin $H_1(\vec{E})$ corresponding to the updated \vec{E}^{cur} is incremented. After N_s transitions, a new, rebaised estimate of $p \equiv p_1$ is generated from $p_1(\vec{E}) = cp_0(\vec{E})H_1(\vec{E})$, where the normalization constant c is normally chosen according to the criterion that p_1 corresponds to a probability distribution. While this procedure initially leads to the Monte Carlo result, in the subsequent iteration, $p_0(\vec{E})$ is replaced by $p_1(\vec{E})$, and the above steps are repeated. The acceptance rule then increases the sampling probability of states with small $p(\vec{E})$ in such a manner that the resulting histogram, $H_n(\vec{E})$, becomes increasingly independent of \vec{E} as the calculation progresses.

3. BIASED MULTICANONICAL SAMPLING

Biasing methods [15,27] increase the multicanonical sampling probability in a limited region \mathcal{R}_k of the system observables. One procedure multiplies the current estimate of $p(\vec{E})$ by a bias enhancement function $F(\vec{E})$ that is constant within \mathcal{R}_k and rapidly increases away from this region, i.e., $\tilde{p}_{m-1}(\vec{E}) = p_{m-1}(\vec{E})F(\vec{E})$ at the beginning of each iteration [13]. This modified transition rule augments the probability of transitions into \mathcal{R}_k ; however, at the end of the iteration loop, an unbiased estimate of p is recovered from $p_m(\vec{E}) = \tilde{p}_{m-1}(\vec{E})H_m(\vec{E})$ as in the importance sampling technique. In the one-dimensional calculations of this paper, \vec{E} corresponds to a single variable E while in the n th multicanonical iteration we employ

$$F(E) = \begin{cases} e^{(E - E_n^L)^2/(2\sigma^2)} & E < E_n^L \\ e^{(E - E_n^R)^2/(2\sigma^2)} & E > E_n^R \\ 1 & E_n^L \leq E \leq E_n^R \end{cases}, \quad (1)$$

where E_n^L and E_n^R are the left and right limits of the region over which $F(E)$ is constant.

If the relative normalization is known, we can combine the estimates, $\hat{p}_i(\vec{E})$, $i=1, \dots, (m-1)$, of $p(\vec{E})$, obtained from either $m-1$ successive iterations or the converged results of $m-1$ biased multicanonical calculations with different bias functions, $F_i(\vec{E})$, for which the corresponding regions \mathcal{R}_i differ. For example, according to the standard importance sampling prescription, the result in the k th histogram bin is obtained by weighing each contribution according to the relative number of samples, $H_i(\vec{E}_k)$, in this bin; that is [14,28],

$$p(\vec{E}_k) = \frac{\sum_{i=1}^{m-1} H_i(\vec{E}_k) \hat{p}_i(\vec{E}_k)}{\sum_{i=1}^{m-1} H_i(\vec{E}_k)}. \quad (2)$$

Unfortunately, the relative normalization of the individual histograms is generally undetermined unless the biasing is sufficiently weak that all calculations yield acceptable estimates of the large-amplitude regions of the pdf. However, if the regions \mathcal{R}_i and \mathcal{R}_{i-1} partially overlap, the respective pdfs \hat{p}_i and \hat{p}_{i-1} differ in the common region by at most a normalization constant if statistical fluctuations are neglected. Below, we present a slight modification of the procedure of [21–23]. The resulting algorithm determines the ratio of the unknown normalization constants by minimizing the MSE of the pdf estimate.

We, accordingly, consider a slightly more general expression,

$$\hat{p}(\vec{E}_k) = \sum_{i=1}^{m-1} A_i w_i(\vec{E}_k) \hat{p}_i(\vec{E}_k), \quad (3)$$

for the pdf estimate at the start of the m th multicanonical iteration, formed by the weighted sum of $m-1$ biased estimates, $\hat{p}_i(\vec{E}_k)$, with

$$\sum_{i=1}^{m-1} w_i(\vec{E}_k) = 1 \quad (4)$$

for all k . The normalization constant, A_i , is defined by

$$p_i(\vec{E}) = \begin{cases} p(\vec{E})/A_i & \vec{E} \in \mathcal{R}_i \\ 0 & \text{otherwise} \end{cases}, \quad (5)$$

in which $p(\vec{E})$ and $p_i(\vec{E})$ are the exact and the biased pdfs, respectively. That is, A_i denotes the integral of $p(\vec{E})$ over the i th biasing region, \mathcal{R}_i . The optimal weight functions can therefore be determined by minimizing the MSE [29] of $\hat{p}(\vec{E}_k)$,

$$\begin{aligned} \text{MSE}\{\hat{p}(\vec{E}_k)\} &\equiv \text{E}\{[\hat{p}(\vec{E}_k) - p(\vec{E}_k)]^2\} \\ &= \sum_{i=1}^{m-1} A_i^2 w_i^2(\vec{E}_k) \text{Var}\{\hat{p}_i(\vec{E}_k)\} + [\text{E}\{\hat{p}(\vec{E}_k)\} \\ &\quad - p(\vec{E}_k)]^2, \end{aligned} \quad (6)$$

with respect to $w_i(\vec{E}_k)$, while simultaneously imposing the normalization condition of Eq. (4) [21,22,28,30]. Here $\text{E}\{\dots\}$ and $\text{Var}\{\dots\}$ represent the expectation and variance over a large number of multicanonical simulations.

To minimize Eq. (6), the constraint of Eq. (4) is introduced through a Lagrange multiplier, λ , and each partial derivative of

$$\hat{f}(\vec{E}_k) = \text{MSE}\{\hat{p}(\vec{E}_k)\} - \lambda \left\{ \sum_{i=1}^{m-1} w_i(\vec{E}_k) - 1 \right\}, \quad (7)$$

with respect to $w_j(\vec{E}_k)$, is subsequently equated to zero. This yields for the optimum weight functions

$$w_i(\vec{E}_k) = \frac{I_i^{(k)} / [A_i^2 \text{Var}\{\hat{p}_i(\vec{E}_k)\}]}{\sum_{j=1} I_j^{(k)} / [A_j^2 \text{Var}\{\hat{p}_j(\vec{E}_k)\}]} \quad (8)$$

The expression $I_i^{(k)}$ is zero if $\hat{p}_i(\vec{E}_k)=0$ and one otherwise, which ensures that the pdf estimate is correctly weighted in the absence of histogram samples.

The variance appearing in Eq. (8) is given in terms of the n_i samples recorded in H_i by [22,29,31]

$$\text{Var}\{\hat{p}_i(\vec{E}_k)\} = \frac{1}{n_i \Delta V} g_i(\vec{E}_k) p_i(\vec{E}_k). \quad (9)$$

Here $g_i(\vec{E}_k)$ should incorporate the covariance between successive samples in the k th histogram bin during the i th multicanonical iteration [22]. For simplicity, however, we approximate $g_i(\vec{E}_k)$ by a constant for all i [22] and further replace $p_i(\vec{E}_k) \rightarrow p(\vec{E}_k)/A_i$ according to Eq. (5). This yields the approximation

$$\hat{p}(\vec{E}_k) \approx \frac{\sum_{i=1}^{m-1} n_i \hat{p}_i(\vec{E}_k)}{\sum_{j=1} I_j^{(k)} n_j / A_j}. \quad (10)$$

We proceed as in [21,22] by observing that

$$A_n = \int_{\mathcal{R}_n} p(\vec{E}) d\vec{E} \approx \sum_{\{k_n | \vec{E}_{k_n} \in \mathcal{R}_n\}} \hat{p}(\vec{E}_{k_n}) \Delta V. \quad (11)$$

Applying Eq. (10) then yields

$$A_n \propto \sum_{\{k_n | \vec{E}_{k_n} \in \mathcal{R}_n\}} \frac{\sum_{i=1}^{m-1} n_i \hat{p}_i(\vec{E}_{k_n})}{\sum_{j=1} I_j^{(k_n)} n_j / A_j}. \quad (12)$$

To solve the above expression, all normalization constants A_n are initialized to unity and then inserted into Eq. (12), which provides new estimates of A_n . The procedure is then iterated until the relative change in A_n is less than a specified value. To avoid numerical divergences, A_n is typically replaced by cA_n after each iteration. In our calculations, the constant c is determined from the condition that the components of the pdf \hat{p} , Eq. (10), sum to unity.

If the intersection of \mathcal{R}_{n-1} and \mathcal{R}_n is disjoint from any other biasing region for every n , the converged A_n/A_{n-1} equals the ratio of the sums of the histogram values within the overlapping region. If, on the other hand, more than one successive biasing region overlaps, the full iterative solution of Eq. (12) is required.

The number of iterations can be considerably reduced if the initial estimates for A_n in Eq. (12) are chosen to minimize the vertical offset between \hat{p}_{n-1} and \hat{p}_n . That is, if $A_n^{(0)}$ represents the initial value of A_n , we minimize the weighted least-squares error [16] with weighting constants w_k^{ls} , namely,

$$\sum_{k=1}^N w_k^{ls} [\hat{p}_{n-1}(\vec{E}_k) - x_n \hat{p}_n(\vec{E}_k)]^2, \quad (13)$$

with respect to $x_n \equiv A_n^{(0)}/A_{n-1}^{(0)}$, $n=2, \dots, (m-1)$. This yields

$$\frac{A_n^{(0)}}{A_{n-1}^{(0)}} = \frac{\sum_{k=1}^N w_k^{ls} \hat{p}_{n-1}(\vec{E}_k) \hat{p}_n(\vec{E}_k)}{\sum_{k=1}^N w_k^{ls} [\hat{p}_n(\vec{E}_k)]^2}. \quad (14)$$

Similarly, $A_n^{(0)}$ can be recursively determined from Eq. (14) after setting $A_1^{(0)}=1$. The choice of weights, w_k^{ls} , is not unique; in our calculations, however, we set w_k^{ls} to the product $(H_n(\vec{E}_k)-1)(H_{n-1}(\vec{E}_k)-1)$ (recall that all histogram bins are initialized to unity).

Accordingly, to determine the pdf of an observable, E , we first compute a Monte Carlo estimate of the pdf, \hat{p}_1 , and select as the initial biasing region an interval centered on the last point, E_0 , in the calculation for which ten or more samples have been recorded. This criterion reflects the empirical observation that, with our computational parameters, these points are normally surrounded by a region of histogram bins that contain at least one sample. The width of the biasing range, ΔE , is then selected for every iteration according to

$$\Delta E = 2 \frac{E_{\max} - E_0}{1 + 2\gamma(N_I - 1)}, \quad (15)$$

where E_{\max} and N_I denote the maximum value of interest of the system observable and the number of biased multicanonical iterations, respectively, and γ is the fraction of the window width by which the biasing range is shifted to the right for each multicanonical iteration, i.e.,

$$E_n^L = E_0 + \Delta E[(n-1)\gamma - 1/2],$$

$$E_n^R = E_0 + \Delta E[(n-1)\gamma + 1/2], \quad (16)$$

for $n=1, \dots, N_I$. That is, $\gamma=0.75$ corresponds to a 25% overlap of the new biasing region with the previous region. Employing N_s samples and $F(E)$ given by Eq. (1), we calculate a biased multicanonical histogram, H_2 , as in Section 2, leading to a new pdf estimate \hat{p}_2 . The ratio of the normalization constants A_1 and A_2 for this first iteration is then determined through Eq. (12), after which the updated pdf is computed from Eq. (10). The biasing region is then shifted to the right according to Eqs. (16), and the procedure is repeated such that the unnormalized pdf estimates $\hat{p}_1, \dots, \hat{p}_{m-1}$, after $m-1$ biased multicanonical iterations, are combined in an analogous fashion to yield the improved pdf estimate for the m th iteration.

Note that our method differs substantially from, e.g., umbrella sampling, in which intermediate results are combined only at the end of the calculation. Further, as noted in [32], our approach enables regions of state space with physically interesting properties to be located and examined dynamically by positioning the biasing region according to the results of previous iterations.

4. NUMERICAL RESULTS

Before establishing the experimental significance of our formalism, we first demonstrate its accuracy through numerical simulations. These will analyze the test case for which the system output variable $E \equiv \tau(\vec{\alpha})$ is the DGD, τ , of a simulated optical fiber emulator. In the calculations of this paper, the emulator consists of $N_{\text{sec}} = 100$ sections of randomly oriented, polarization-maintaining (PM) fiber [33]. The DGD of a single fiber section, τ_{sec} , is determined by the requirement that the average DGD $\tau_{\text{avg}} = \tau_{\text{sec}} \sqrt{8N_{\text{sec}}/3\pi} = 25$ ps. The system parameters, $\vec{\alpha}$, then correspond to the coupling angles between the PM fiber segments [6]. Our calculations further employ ten 2×10^4 -sample iterations with $\gamma = 0.3$. Events are recorded in 100 equal-size histogram bins in the interval $0 < \tau/\tau_{\text{avg}} < 10$.

We first examine the rate of convergence afforded by Eq. (12) in Fig. 1, which displays the maximum relative error,

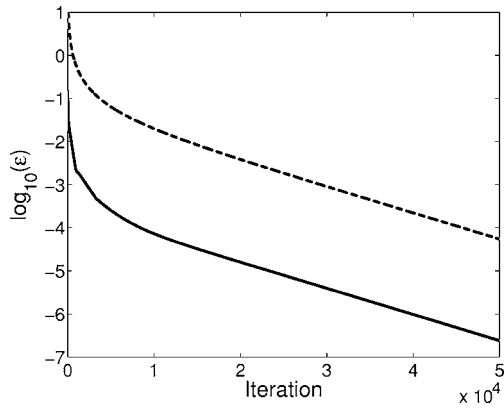


Fig. 1. Maximum relative error, Eq. (17), for the initial estimate of the normalization constants of Eq. (14) (solid curve) and for the iterated results $A_n^{(0)} = 1$, $n = 1, \dots, N_I$ for these constants (dotted curve). Results are illustrated for a 100-segment fiber emulator with $\tau_{\text{avg}} = 25$ ps and ten biased multicanonical iterations with 2×10^4 points per iteration with $\gamma = 0.3$.

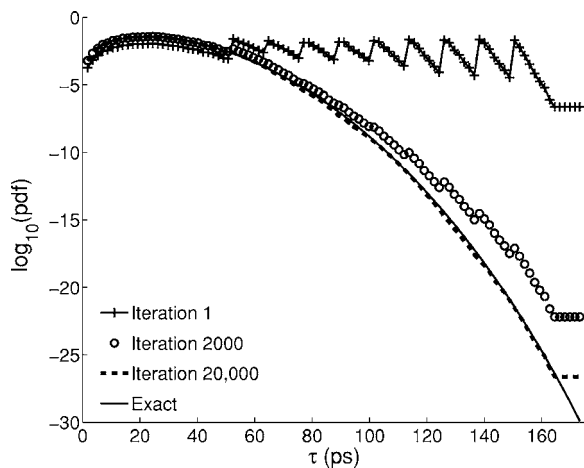


Fig. 2. Pdf, Eq. (10), obtained from the calculation of Fig. 1 after joining the results of ten different biased calculations after 1 (crosses), 2000 (circles), and 20,000 (dotted curve) iterations of Eq. (12). The solid curve is the exact result of [8]. The normalization constants are here initialized to $A_n^{(0)} = 1$, $n = 1, \dots, N_I$.

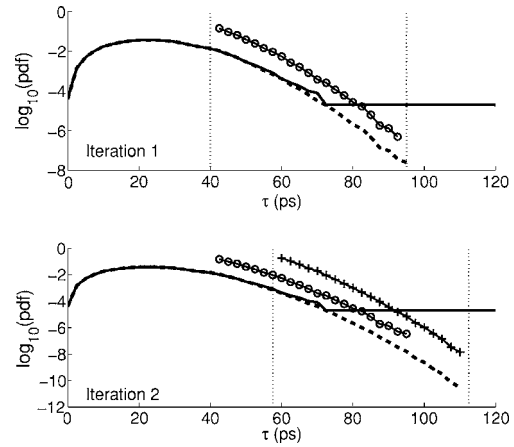


Fig. 3. Top, biased pdf estimate \hat{p}_2 before (circles) and after (dashed curve) combination with the initial Monte Carlo estimate \hat{p}_1 (solid curve) according to Eq. (10) with $\gamma = 0.3$. Bottom, analogous results for the second biasing iteration in which \hat{p}_1 , \hat{p}_2 , and \hat{p}_3 (crosses) are combined according to Eq. (10) (solid curve). The vertical lines indicate the biasing region.

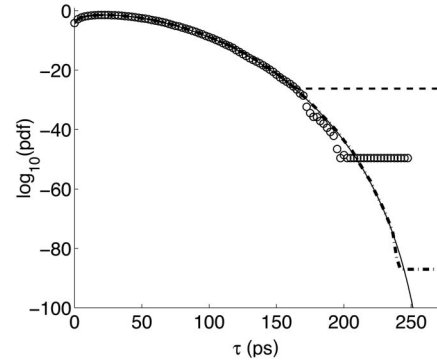


Fig. 4. Pdf of the DGD of a 100-section fiber emulator calculated with fifty 5×10^4 -sample (dashed-dotted curve) and fifteen 1.2×10^4 -sample (dashed curve) iterations of the biased multicanonical method, with the standard multicanonical procedure for fifteen 1.67×10^5 -sample iterations (circles) and with the analytic result (solid curve).

$$\epsilon = \max_n \left[\frac{A_n^{(i)} - A_n^{(i-1)}}{A_n^{(i-1)}} \right] \times 100\%, \quad (17)$$

of the normalization constants, after the i th iteration, $i = 2, 3, \dots, 50,000$. The initial estimates $A_n^{(0)} = 1$ yield the dotted curve in Fig. 1, while the initial estimates of Eq. (14) instead yield the solid curve in the figure. The combined pdf estimates obtained from Eq. (10) after 1 (crosses), 2000 (circles), and 20,000 (dotted curve) iterations of Eq. (12) for the case in which all A_n were initialized to unity are presented in Fig. 2.

Next, in Fig. 3 we illustrate the combination of piecewise biased estimates of the pdf. The solid curve in the top graph of Fig. 3 is the initial pdf estimate, \hat{p}_1 , obtained from a Monte Carlo simulation, while the subsequent biased estimates, \hat{p}_2 , before and after the application of Eq. (12), are indicated by circular markers and dashed curves, respectively. The quantity \hat{p}_2 is displayed only within the region of constant bias, as the histogram values outside this region do not affect the updated pdf esti-

mate in our procedure. The second graph in Fig. 3 depicts the corresponding results for the second iteration for which the dotted curve is a combination of the three results \hat{p}_1 , \hat{p}_2 , and \hat{p}_3 according to Eq. (10).

After fifty 5×10^4 -sample iterations with $\gamma=0.25$, we generated the dashed-dotted curve in Fig. 4 with our biased multicanonical method, which is compared with the analytic result (solid curve) [8]. The standard multicanonical procedure with fifteen 1.67×10^5 -sample iterations, which was found empirically to yield near-ideal results for the pdf in the low-probability tail region, instead yields the curve designated by circular markers. Finally, the dashed curve of Fig. 4, obtained after fifteen 1.2×10^4 -sample iterations with $\gamma=0.25$, exhibits a level of accuracy similar to the standard multicanonical procedure. Evidently, our biased multicanonical method, even without optimization of computational parameters, yields a considerable increase in computational efficiency compared with the standard multicanonical procedure with the same number of samples.

5. EXPERIMENTAL MEASUREMENTS

To demonstrate that our procedure can be applied experimentally, we measured the pdf of the DGD, τ , of an eight-stage optical fiber emulator, following the procedure of [12]. The experimental setup, see Fig. 5, employed eight General Photonics PolarRITE II three-axis polarization controllers separated by unequal-length PM fibers. As in Section 4, $E \equiv \tau(\vec{a})$ represents the relevant system observable, namely, the DGD, while \vec{a} corresponds to the 24 input voltages applied to the eight polarization controllers. The DGD of the fiber emulator was measured by applying the Jones matrix eigenanalysis (JME) procedure to the output of an HP 8509B polarization analyzer combined with a Tunicas PRI tunable laser [34]. The mean and maximum DGDs for the system are approximately 11.7 ps and 37.4 ps, respectively.

In our experiment, the Jones matrix was measured at two wavelengths displaced by 0.1 nm by cycling an EO-Space eight-stage polarization controller through five random input polarization states and reading the corresponding output Stokes vectors from the analog outputs of the HP 8509B with a National Instruments PCI-6221 input card. The Jones matrix was computed from the output values according to the least-squares method of [35,36], at a rate five times faster than the native HP

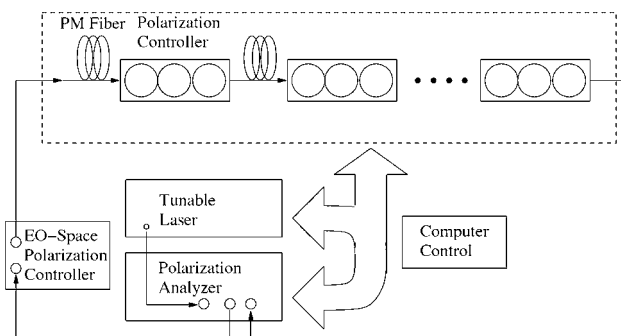


Fig. 5. Biased multicanonical experimental setup.

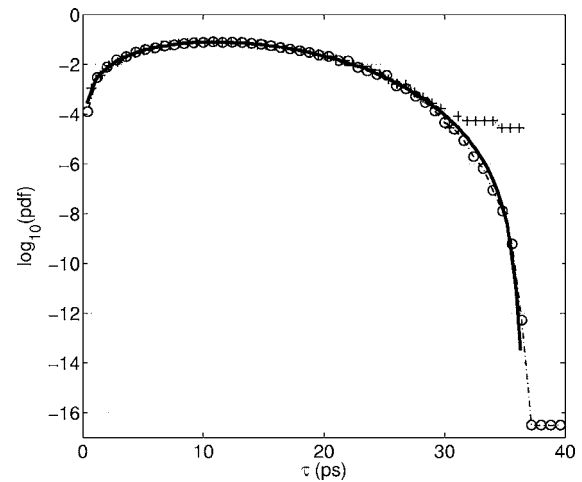


Fig. 6. Experimentally determined pdf of the DGD of an eight-section fiber emulator for 45,000 samples measured with the piecewise biased multicanonical method (circles) and the standard Monte Carlo procedure (crosses). The solid curve indicates the corresponding numerical results for three 5×10^5 -sample iterations in the standard multicanonical method.

8509B JME measurement algorithms, although further improvements could be attained with relatively minor component modifications.

Employing 10,000 Monte Carlo measurements followed by seven 5,000-sample biased multicanonical iterations with $\gamma=0.5$, the piecewise procedure of Section 3 yielded the circular markers in Fig. 6 for the measured pdf of the DGD, τ , of the fiber emulator. For comparison, the points marked as crosses in the figure were instead generated with a standard 45,000-point Monte Carlo measurement. Finally, the solid curve of Fig. 6 indicates the simulated pdf for three 5×10^5 -sample iterations of the standard multicanonical algorithm in which the DGD for each PM fiber segment corresponded to the values measured in the experimental setup. Although the number of multicanonical iterations and sampling events were chosen to produce probabilities of occurrence comparable with the experimental results, the agreement between the calculated and the experimentally determined pdfs is still notable [9,33]. Evidently, our aligned piecewise biasing procedure computes the low-probability regions of the pdf with considerably greater efficiency than the standard multicanonical method.

6. CONCLUSIONS

We have employed an iterative method to combine the intermediate results of strongly biased multicanonical calculations. Our procedure can be immediately adapted to any biasing formalism, such as directly calculating quantities other than the pdf [4], raising the intermediate pdf or histogram variable to a power (power method), dynamically biasing in multidimensional space to locate regions with physically interesting properties [32], and employing highly confining bias functions to strictly limit the statistical events to a small region of parameter space (barrier method) [14]. Since in each case the sampling region is limited to a far smaller region of parameter space than in standard multicanonical calculations, consider-

able increases in computational efficiency are possible, especially in high-dimensional problems, which is of great significance in experimental applications.

ACKNOWLEDGMENTS

The authors thank the National Sciences and Research Council of Canada, Nortel, the Center for Electrophotonic Materials and Devices (CEMD), and the Ontario Research and Development Challenge Fund (ORDCF) for continued financial support.

REFERENCES

1. P. Smith, M. Shafi, and H. Gao, "Quick simulation: a review of importance sampling techniques in communications systems," *IEEE J. Sel. Areas Commun.* **15**, 597–613 (1997).
2. J. Bucklew and R. Radeke, "On the Monte Carlo simulation of digital communication systems in Gaussian noise," *IEEE Trans. Commun.* **51**, 267–274 (2003).
3. B. Berg, "Introduction to multicanonical Monte Carlo simulations," *Fields Inst. Commun.* **26**, 1–24 (2000).
4. D. Yevick, *A First Course in Computational Physics and Object-Oriented Programming with C++* (Cambridge U. Press, 2005).
5. D. Yevick, "The accuracy of multicanonical system models," *IEEE Photon. Technol. Lett.* **15**, 224–226 (2003).
6. D. Yevick, "Multicanonical communication system modeling—application to PMD statistics," *IEEE Photon. Technol. Lett.* **14**, 1512–1514 (2002).
7. J. Gordon and H. Kogelnik, "PMD fundamentals: polarization mode dispersion in optical fibers," *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4541–4550 (2000).
8. M. Karlsson, "Probability density functions of the differential group delay in optical fiber communication systems," *J. Lightwave Technol.* **19**, 324–331 (2001).
9. G. Foschini and C. Poole, "Statistical theory of polarization dispersion in single mode fibers," *J. Lightwave Technol.* **9**, 1439–1456 (1991).
10. M. Jeruchim, P. Hahn, K. Smyntek, and R. Ray, "An experimental investigation of conventional and efficient importance sampling," *IEEE Trans. Commun.* **37**, 578–587 (1989).
11. T. Lu, D. Yevick, L. Yan, B. Zhang, and A. E. Willner, "An experimental approach to multicanonical sampling," *IEEE Photon. Technol. Lett.* **16**, 1978–1980 (2005).
12. T. Lu, D. Yevick, B. Hamilton, D. Dumas, and M. Reimer, "An experimental realization of biased multicanonical sampling," *IEEE Photon. Technol. Lett.* **17**, 1583–1585 (2005).
13. T. Lu and D. Yevick, "Biased multicanonical sampling," *IEEE Photon. Technol. Lett.* **17**, 1420–1422 (2005).
14. D. Yevick and T. Lu, "Improved multicanonical algorithms," *J. Opt. Soc. Am. A* **23**, 2912–2918 (2006).
15. R. Wolfe, M. Jeruchim, and P. Hahn, "On optimum and suboptimum biasing procedures for importance sampling in communication simulation," *IEEE Trans. Commun.* **38**, 639–647 (1990).
16. N. Alves, B. Berg, and R. Villanova, "Ising-model Monte Carlo simulations: density of states and mass gap," *Phys. Rev. B* **41**, 383–394 (1990).
17. P. Virnau and M. Muller, "Calculation of free energy through successive umbrella sampling," *J. Chem. Phys.* **120**, 10925–10930 (2004).
18. G. Torrie and J. Valleau, "Monte Carlo free energy estimates using non-Boltzmann sampling: application to the sub-critical Lennard–Jones fluid," *Chem. Phys. Lett.* **28**, 578–581 (1974).
19. G. Torrie and J. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling," *J. Comput. Phys.* **23**, 187–199 (1977).
20. C. Bennett, "Efficient estimation of free energy differences from Monte Carlo data," *J. Comput. Phys.* **22**, 245–268 (1976).
21. S. Kumar, D. Bouzids, R. Swendsen, P. Kollman, and J. Roseberg, "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method," *J. Comput. Chem.* **13**, 1011–1021 (1992).
22. M. Souaille and B. Roux, "Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations," *Comput. Phys. Commun.* **135**, 40–57 (2001).
23. A. Ferrenberg, D. Landau, and R. Swendsen, "Statistical errors in histogram reweighting," *Phys. Rev. E* **51**, 5092–5100 (1995).
24. G. Smith and A. Bruce, "A study of the multi-canonical Monte Carlo method," *J. Phys. A* **28**, 6623–6643 (1995).
25. B. Berg, "Algorithmic aspects of multicanonical simulations," *Nucl. Phys. B* **63A-C**, 982–984 (1998).
26. B. Berg, "Multicanonical recursions," *J. Stat. Phys.* **82**, 323–342 (1996).
27. N. Mandayam and B. Aazhang, "Importance sampling for analysis of direct detection optical communication systems," *IEEE Trans. Commun.* **43**, 229–239 (1995).
28. E. Veach, "Robust Monte Carlo methods for light transport simulation," Ph.D. thesis (Stanford University, 1997).
29. D. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization* (Wiley, 1992).
30. A. Acharya, "Free energy differences: representations, estimators and sampling strategies," Ph.D. thesis (University of Edinburgh, 2004).
31. A. Ferrenberg and R. Swendsen, "Optimized Monte Carlo data analysis," *Phys. Rev. Lett.* **63**, 1195–1198 (1989).
32. T. Lu, W. Huang, D. Yevick, M. O'Sullivan, and M. Reimer, "Multicanonical comparison of polarization-mode dispersion compensator performance," *J. Opt. Soc. Am. A* **22**, 2804–2809 (2005).
33. A. Djupsjobacka, "On differential group-delay statistics for polarization-mode dispersion emulators," *J. Lightwave Technol.* **19**, 285–290 (2001).
34. B. Heffner, "Automated measurement of polarization mode dispersion using Jones matrix eigenanalysis," *IEEE Photon. Technol. Lett.* **4**, 1066–1068 (1992).
35. M. Reimer and D. Yevick, "Least-squares analysis of the Mueller matrix," *Opt. Lett.* **31**, 2399–2401 (2006).
36. D. Sandel, V. Mirvoda, S. Bhandare, F. Wust, and R. Noe, "Some enabling techniques for polarization mode dispersion compensation," *J. Lightwave Technol.* **21**, 1198–1210 (2003).